# Literature Review of Different Machine Learning Algorithms for Credit Card Fraud Detection

Nayan Uchhana, Ravi Ranjan, Shashank Sharma, Deepak Agrawal, Anurag Punde

*Abstract: Every year fraud cost generated in the economy is more than $4 trillion internationally. This is unsurprising, as the return on investment for fraud can be massive. Cybercrime specialists estimate that an investment of 1 million dollars into fraud or attack can net up to $100 million. Financial institutions such as commercial and investment banking operations are increasingly being targeted. And we know that the only way to fight fraud effectively is through the use of advanced technology. The answer lies in relying on advanced analytics and enterprise-wide data storage capabilities that support the use of artificial intelligence (AI) and machine learning (ML) approaches to stay one step ahead of criminals. AI is best suited to defend against today's fast-changing and complex bank fraud, where new threats are under development every day. Approaches relying on fragmented and siloed data, rules-based approaches or traditional point-solutions are no longer acceptable. These approaches are not only ineffective, but they are extremely costly to banks and financial services firms because they force legal and compliance teams to spend a lot of time trying to gain access to the data they need. By relying on advanced analytics and AI and ML capabilities, fraud and compliance units can spend their time working on more-complex fraud issues. Manual investigation can be reduced through the use of complex algorithms powered by ML, often in conjunction with rules, a combination that offers significant advantages over purely based -rules fraud detection. In this paper, we have included different machine learning algorithms used to detect credit card frauds and also provide a comparative study between different algorithms.*

*Keywords- Machine Learning, Credit Card Fraud Detection, KNN, Clustering*

## I. INTRODUCTION

At present time were worlds are growing very fast as compared to past years. The biggest reason behind the fast growing is trading. Trading through internet is said to be e – commerce, where goods get exchanged through online services. Basically, e – Commerce is a platform where people get electronic URL as their shop for purchasing goods and items to buy or pay through internet. For buying they need to pay money with that they pay through different services available most probably through credit cards or debit cards with these two cards we can pay money to the trader but when cards come into the play it brings a new term with it that is cyber security.

**Nayan Uchhana,** Pursuing B.E. (CSE), Indore Institute of Science & Technology Indore, India.

**Ravi Ranjan,** Pursuing B.E. (CSE), Indore Institute of Science & Technology Indore, India.

**Shashank Sharma,** Software Developer, Bachelors of Engineering in Information Technology, IET DAVV, Indore, India.

**Deepak Agrawal*,** Assistant Professor, Department of Computer science, IIST, Indore, India.

**Anurag Punde,** Assistant Professor, Department of Computer science, AITR, Indore, India.

These are a type of lock that keeps online transaction safe from any digital attack. The digital attack can take place in many ways but to attack the fraudster need a to find loop hole through which they can get the access once they found the hole the account will have slightly suspicious behavior that is sent to the next software that is security information and event management system software now IRP (interim resolution professional) is used to prevent a serious security disaster. Mostly the online transaction takes place under banks operations through credit card or debit card. There are many different types of credit card frauds like.

❖ stolen credit card

❖ Use of card that is through information fraudsters doesn't require the physical card to access and make transaction through account.

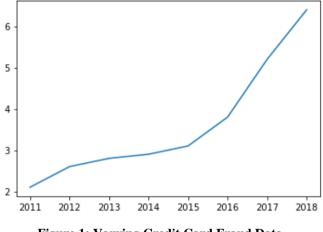❖ account takeover

❖ Misplace card.



**Figure 1: Varying Credit Card Fraud Data**

Now if we have look over the credit card fraud report in US. It has shown a great increment in the growth of frauds.

The above result are generated by CVV (card verification value). The CVV is a 3 or 4 line code present on credit cards and debit cards .we use CVV number to make transactions online .the details of debit or credit cards were required to procced the transaction here you don't need the password your card details are enough to make transactions online so, to stop it we use the following techniques :

❖We use the device pattern or device fingureprint to approve/decline the transactions .

❖Also we can search the blacklist details of the fraud card if already avaliable we don't approve it .

101

❖ We can use the tools to monitoring the anomalies that helps to identify the transaction which doesn't have certain rules.

❖ You can set a flag transaction of your account that may not allow the fraudster to make a large transaction.that generally happpens.

❖ By limiting the attemp of transactions a customer can protective layer that work like a threshold where fraudster cannot use card number software generater to hit and trial it is a velocity check.

Now, if we focus on the figure-1 given above from year 2011 – 2018 there is increment in the fraud losses. Even after all the precautions we take still they find a loop hole to get into the system and crack it. In present world investigators are not able to check all transactions even the fraud detection system monitors all the transactions and verifies all the activities take place with the account and its details [A Review on Credit Card Fraud Detection Using Machine Learning]. Every year there is a big difference in the growth of frauds even after having many tools and techniques but still fraud takes place in a huge amount. Now, if we have a look on UK's economy where UK is the world's third largest e – Commerce market that is going to be $314 billion in 2021 where most of the people use online payment methods for their purchase that indirectly attract fraudsters. Since the use of cash payment in UK is 44 percent whereas card payment is 56 percent and the fraud on retailer websites is cost over $400 million also fraud increased by 25 percent. Similarly, every country is facing the same problem with credit card frauds so to deal with it, we are going to approach the latest technologies that is machine learning (ML) and artificial intelligence (AI). Below given is the classification of the credit card fraud detection based on supervised and unsupervised learning systems. Which were further divided. For supervised learning we have AIS, EXPERT SYSTEM, ANN(BP)and so on. For unsupervised we have HMM, ANN(SOM), FUZZY SYSTEM. All the techniques need a data set to examine and learn from it with high efficiency so for different datasets we required different techniques that can answer in more accurate way [A Survey of Credit Card Fraud Detection Techniques: Data and Technique Oriented Perspective].And the most popular technologies are Artificial Neural Network , Fuzzy Logic, Genetic Algorithm, Logistic Regression Decision tree, Support Vector Machines, Bayesian Networks,Hidden Markov Model, K-Nearest Neighbor .
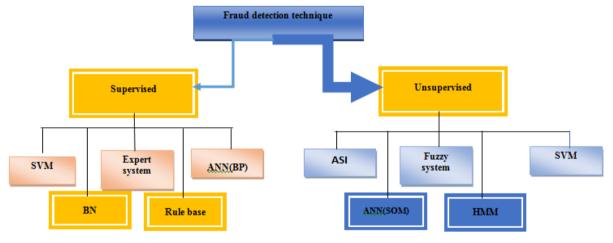


**Figure 2: Machine Learning Techniques for Credit Card Fraud Detection**

By using these two technologies, we can solve the issue of credit card frauds up to a good extent. Mostly fraudster's fraud by hacking your data of the card.They doesn't need a physical card to hack it. They can easily make transactions with the card details. There isn't any standard method to stop it from root cause but can identify by using some methods. So, with the use of machine learning algorithms we can train the model and predict the outcome of the transaction by feeding the model with credit card fraud data and using supervised learning to classify the categories of fraud and secrecy.

## II. LITERATURE REVIEW

**Rimpal R. Popat with Jayesh Chaudhary:** They made a survey on credit card fraud detection, considering the major areas of credit card fraud detection that are bank fraud, corporate fraud, Insurance fraud. With these they have focused on the two ways of credit card transactions i) Virtually (card, not present) ii) With Card or physically present. They had focused on the techniques which are Regression, classification, Logistic regression, Support vector machine, Neural network, Artificial Immune system, K-nearest Neighbor, Naïve Bayes, Genetic Algorithm, Data mining, Decision Tree, Fuzzy logic-based system, etc. In which, they have explained six data mining approaches as theoretical background that are classification, clustering, prediction, outlier detection, Regression, and visualization. Then have explained about existing techniques based on statistical and computation which is Artificial Immune system (AIS), Bayesian Belief Network, Neural Network, Logistic Regression, Support Vector Machine, Tree, Self-organizing map,Hybrid Methods, As a result, they had concluded that all the present machine learning techniques mentioned above can provide high accuracy for the detection rate and industries are looking forward to finding new methods to increase their profit and reduce the cost. Machine learning can be a good choice for it. [A Survey on Credit Card Fraud Detection using Machine Learning].

**Mohamad Zamini:** purposed an unsupervised fraud detection method using autoencoder based clustering.The autoencoder **is** anauto associatorneural network they have used it to lower the dimensionality, extract the useful features, and increase the efficiency of learning in a neural network. They had used European dataset with 284807 transactions in which 0.17% is the fraud and trained there autoencoder based clustering with the following parameters
Number of iterations = 300

Number of clusters = 2

Clustering initialization = k-means++

Divergence tolerance = 0.001

Learning rate of the model = 0.1

 Number of epochs = 200

Activation function = elu, Relu.

As a result, they got their training loss as 0.024 and validation loss as 0.027 and the mean of not fraud data 75% less than the mean of reconstructive error that is 25% the design of there is model is context-free. In concern about the model predictions, the True positive are 56,257, False-negative is 607, False positive are 18, True negativesis 80 and the best preferred are (56,257 + 80 = 56,337). The right predictions made are 56,337 out of 284807. [Credit Card Fraud Detection using autoencoder based clustering].
**Shiyang Xuan:**they made a comparison based on two random forests.

Random-tree-based random forest

CART-based random forest.

They use different random forest algorithms to train the behavior features of normal and abnormal transactions and both of the algorithms are different in their base classifications and their performance. They applied both of the algorithms on the dataset e-commerce company in China. In which the fraud transaction in the subsets ratio is 1:1 to 10:1. As a result, accuracy from the random-tree-based random forest is 91.96% whereas in CART-based random forest is 96.7%. Since the data used is from the B2C dataset many problems arrived such as unbalanced data. Hence, thealgorithm can be improved. [Random Forest for Credit Card Fraud Detection].

**Dejan Varmedja:**proposed the various machine learning algorithms and analyzed them concerning to credit card fraud detection methods. The various methods of machine learning are Logistic Regression, Naïve Bayes, Random Forest, Multilayer Perceptron. Here for multilayer perceptron (ANN) is used (Artificial neural network) which consist of 4 hidden layers and relu activation functioned is used that is to avoid negative values and optimizer used is Adam for its best performance As a result for Logistic regression the accuracy score is 97.46% with the data set containing 56962 samples in which 98 fraud transactions. For the same dataset Naïve Bayes and Random Forest, accuracy score is 99.23% and 99.96% respectively. At last for ANN it was 99.93% of accuracy as we can observe that random forest gives the best result in case of credit card fraud detection. [ Credit Card Fraud Detection - Machine Learning methods]

**Changjun Jiang:**proposed a novel fraud detection method that has four stages they first utilize the historical transaction data to divide them into groups to form clusters of transactions having the same behavior then thus they came up with a sliding windowstrategy to aggregate transactions.This algorithm is used to characterize the behavioral pattern of a cardholder then after aggregation, we use the new window formed the feature extraction is done. At last, the classification takes place and classifies behavioral patterns and assignments. As a result, their method ofLogistic Regression with raw data (RawLR), Random Forest with aggregation data (AggRF), and Random Forest and feedback technique with aggregation data (AggRF +FB) are the best method with 80% accuracy as compared to other methods.[ Credit Card Fraud Detection: A Novel Approach Using Aggregation Strategy and Feedback Mechanism]

**Sahil Dhankhad:** They applied supervised machine learning algorithms on the real-world data set and then used those algorithms to implement a super classifier using ensemble learning and thenthey compared the performance of supervised algorithms with their implementation of a super classifier. They used ten machine learning algorithms such as Random Forest, Stacking Classifier, XGB Classifier, Gradient Boosting, Logistic Regression, MLP Classifier, SVM, Decision Tree, KNN, Naïve Bayes. And compared the accuracy, Recall Precision, confusion matrix with the result of their super classifier. As a result, they found that the Logistic Regression is better for predicting fraud transactions. [Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study].

**Kuldeep Randhawa:**They used twelve machine learning algorithms for credit card fraud detection in which their range standard from a neural network to deep learning. They are tracing the performance of benchmark and real-world datasets. In addition, the AdaBoost and majority voting methods are applied for forming the hybrid models. As there related study explains about single and hybrid models. For both the parameters (Benchmark and real-world datasets) they had given the results using there twelve selected algorithms that are Naïve Bayes, Random Forest, Decision Tree, Gradient Boosted Tree, Decision Stump, Random Tree, Neural Network, Linear Regression, Deep Learning, Logistic Regression, SVM, Multilayer Perceptron. As a result, when standard algorithms used with AdaBoost and majority voting methods under benchmark data the best accuracy and sensitivity acquired by Random Forest algorithm 95% and 91% respectively. When experimented with real-world data the accuracy rate is still above 90% even with 30% noise in the dataset. MCC (Mathews correlation coefficient) is standard to measure the performance of a model so in case of majority voting the best MCC score is 0.823 whereas 0.942 with 30% of noise added to the dataset. [Credit Card Fraud Detection Using AdaBoost and Majority Voting]

**Sai Kiran:** proposed an improved algorithm for credit card fraud detection. That is named as Naïve Bayes improved K-nearest Neighbor method (NBKNN). They have used a dataset on which they had applied the algorithms to identify the fraudulent transaction in the taken dataset.

The dataset has the record of European Cardholders who made a transaction using their credit cards within 2 days they made 284,807 transitions in which 492 transaction is fraudulent. The techniques used which were used are classification techniques but work differently on the same dataset. They had used both of the techniques (Naïve Bayes and k-nearest neighbor) to enhance the accuracy and flexibility of the algorithm. As a result, they got the accuracy of approximately 95% from Naïve Bayes and 90% from K -nearest Neighbor techniques.

## III. COMPARISON OF DIFFERENT MACHINE LEARNING ALGORITHM

In this study total, five algorithms are used which are SVM (support vector machine), Logistic Regression, Gaussian NB, K-Neighbor classifier, Random Forest. After experimenting with this algorithm, we have applied the Grid Search algorithm to find out the best parameters from the algorithm that gives a good accuracy score to our model.

**SVM:** We have used the SVM algorithm because it gives good results with non-linear classification-based problems also it works with the uneven structure of data and the risk of over fitting in a model is reduced to very less.

**Logistic Regression:** It works best when applied on data which have correlated attributes in it. Its computational resources demand is very less. Since it is easy to implement, we can mark it as our benchmark and then work on other algorithms. It usually has the best informative output for the classification method.

**Gaussian NB (Naïve Bayes):** It is an algorithm based on conditional probability so it good to with real-time data set with it. It can lead to a good formation of recommendation system. It can be applied on large datasets. It uses a Formula to calculate conditional probability. That Formula is:

$$P\,(A|B) = \frac{P\,(B|A) * P\,(A)}{P\,(B)}$$

Where P (A|B) = Posterior Probability, P(B|A) = Prior Probability, P(A) = Likelihood, P(A) = Evidence. As a result, it gives a Probabilistic Prediction with less training dataset. It can also handle the continuous and discrete data.

**K-Neighbor classifier:** It is good at handling the noisy data. It's a memory-based approach in which we can use both the classification types (Binary class and Multi class) that too without any extra efforts. Also, we can use with Classification and Regression both. Parameter selection is hard at first parameter later it gets aligned with the first parameter.

**The Random Forest:** In it data doesn't need to be rescaled or transformed. It can be applied on Classification and Regression problems. The algorithm divides the data based on their feature and each tree has high variance and low bias that leads to a good result. It trains the model with high speed also easy to implement and can handle a good amount of feature loss and errors in data set.

**Table 1: The advantages and disadvantages**

| Models | Advantages | Disadvantages |
|---|---|---|
| **SVM (Support Vector Machine)** | Small change in data will not affect the model stability. | Long training time, lots of memory required. |
| **Logistic Regression** | It tells about coefficient size with direction (positive/negative) | If (No. of observation< features) than it may lead to over fit. |
| **Gaussian Naïve Bayes** | The prediction and training process is fast and gives best result with small data sets. | It assumes that all the attributes in a training data is independent of other features. |
| **K-Neighbor Classification** | Since it doesn't require training time new data can easily added to it. | Need scaled data. It can't work with high dimensionality. |
| **Random Forest** | It uses bagging and ensemble learning as a result it reduces over fitting and improves the accuracy of a model. | It is complex in nature. |

## IV. EXPERIMENT

**Dataset**

In credit card fraud detection, we are using a dataset that have total 284807 records of transactions that occurred in just two days on September 2013 by European cardholders. In which there are 492 were fraud transactions and 284,315 are positive transactions as compared to over all dataset there is 0.172% of fraud transactions took place. It also has 31 features which are confidential because of sensitive features; the hidden features are named from V1 to V28. Just to make data more efficient the data is been transformed by PCA (principal component analysis) as a result the dimensionality got redacted and only numerical values are available as input data. Only features which haven't been transformed by PCA are Time, Amount and the feature Class is our target column in which it represents 1 as fraud transaction and 0 as positive transaction.
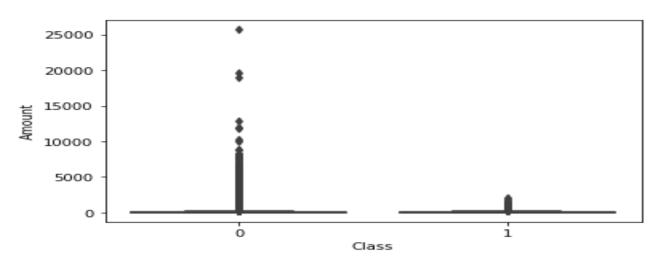
**Figure 3: Here 0 represents the positive transitions and 1 represents the fraud transactions**

**Libraries**

There are four libraries used in the experiment to generate the results. That are:

- ❖ **Pandas:** It is used to read excel and CSV files with that it can be used to organize the data into data frames and operate some functions on it.

- ❖ **Matplotlib:** It is used to visualize the data in 2 dimensions in the form of graphs. We can form many different types of graphs by using it and adjust the size and color of the graphs.

- ❖ **Seaborn:** It is also used to form graphs and visualization of data but the graphs formed by it is slightly different than Matplotlib it has a different type of graphs like boxplot, heatmap, etc.

- ❖ **Scikit-learn:** It is used to import the class of algorithms like SVC, KNN, Regression, Classifiers, etc.

**Implementation of algorithms**

When we apply SVM (support vector algorithm) to the dataset we split it into training data and testing data by applying a parameter to our train-test-split method. As parameter, we mention the test size that is equal to 0.25 for the dataset. That means the overall data will be divided into two parts in which for training purposes it will 0.75 and the rest will be for the testing purposes that is 0.25 as mentioned. Also, we pass all the attributes and target class in two different variables. After these, we form a model of our SVC method in which training data is fitted to train the model. Once the model is ready it predicts the values for our testing data. As a result, we get the classification report and confusion matrix.
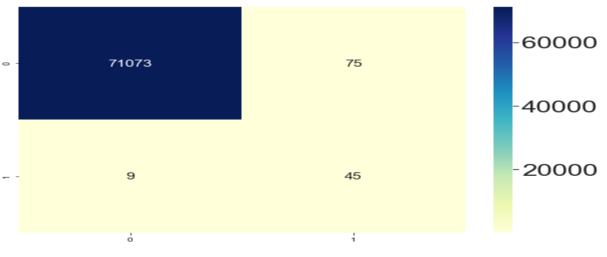


**Figure 4: confusion matrix of SVM**

The above matrix is a result of the SVM algorithm which has True positive is 71073, false negative is 75, false positive is 9, True negative is 45. In this model the right predicted values according to our confusion matrix are 71,118 and wrong predicted values are 84. When we use the standard performance measurement score that is MCC (Matthews Correlation coefficient) for binary Classification, the best score for SVC is 0. 558.

The process for all other algorithms is the same but the methods are changed. Likewise, for Logistic Regression, Naïve Bayes, K-nearest neighbor and Random Forest used methods are Logistic Regression, Gaussian NB, K-Neighbors Classifier, and Random Forest Classifier respectively. As a result, we have generated the Confusion matrix for each of the algorithms and on the best resultant algorithm.
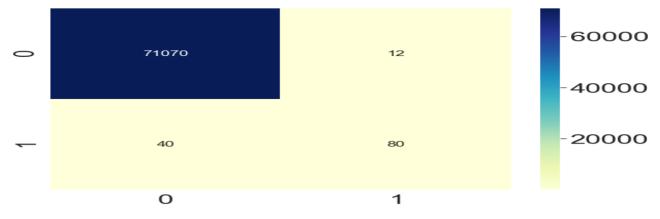
**Figure 5: confusion Matrix of Logistic Regression and Naïve Bayes**

The above Matrix is a result of Logistic Regression. In which True Positive is 71070, False positive is 40, False negative is 12, True negative is 80. In this model, the right predictive values are 71,150 and wrong predictive values are 52. And the MCC score for the Logistic Regression model is 0.761. since the best score of MCC is 1 as compared to MCC's highest score it is an improved algorithm for credit card fraud detection.

The matrix formed by NaïveBayes algorithm is the same as Logistic Regression for this dataset. That means for the taken dataset the MCC score by Naïve Bayes algorithm is 0.761. The algorithms may give different results for another credit card fraud detection dataset that may vary in size because of the f1 score and recall values still different in both the algorithms.
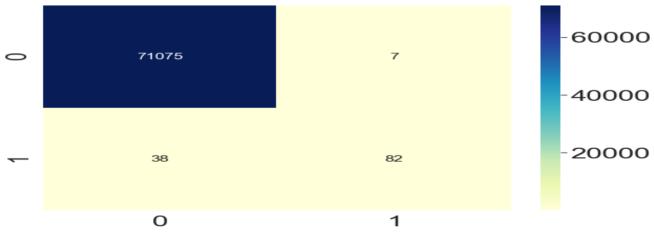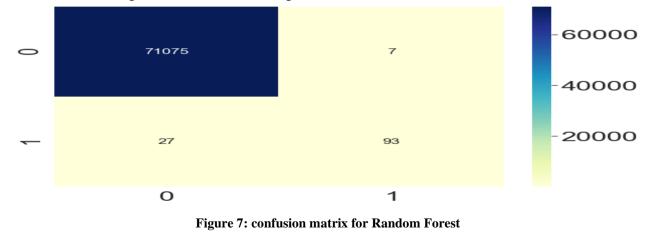


**Figure 6: confusion matrix for K nearest neighbor**

The above-shown matrix is a result of the KNN algorithm. In which the True positive values are 71075, False positive values are 38, False-negative value is 7, True negative values are 82, For this model, the right predicted values are 71,157, and the wrong predicted values are 45. And the MCC score for this model of KNN is 0.793.



**Figure 7: confusion matrix for Random Forest**

The above-shown matrix is a result of the Random Forest algorithm. In which the parameters used are n_estimators are 190, the criterion used is "entropy", max_depth is 10. For which the True Positive values are 71075, False positive are 27, False-negative values are 7 and False-positive values are 38. The right predictions made by this algorithm are 71,168 and wrong predictionsare 34. At last, the MCC score of Random Forest with the above-mentioned parameters is 0.848.

**Table2. For comparing the models MCC values**

| Models | Recall value | F1-score | MCC value |
|---|---|---|---|
| SVM | 0.92 | 1.00 | 0.558 |
| Naïve Bayes | 0.91 | 0.98 | 0.761 |
| Logistic Regression | 0.83 | 1.00 | 0.761 |
| KNN | 0.84 | 1.00 | 0.793 |
| Random Forest | 0.89 | 1.00 | 0.848 |

Now if we have a look at the comparative table so the best score for MCC is 0.848 that is given by Random forest with the random parameters. Then we picked the Random Forest algorithm and apply the GridSearch method to find out the best Parameters and again with the new parameters we have generated a model and compared the resultants.
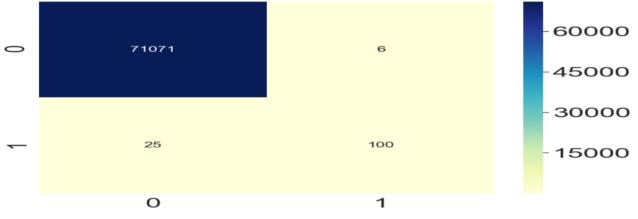


**Figure-8 confusion matrix of Random forest with Grid Search parameters**

The above-generated matrix is a result of the Random forest with the Grid Search parameters and the parameters are n_estimators are 500, max-features used is auto, max_depth is 10, the criterion is entropy. As a result of the confusion matrix, it has 71071 True positive values, 6 false negative values, 25 False positive values, and 100 True negative values. That states that right predictive values are 71,171 and wrong predictive values are 31. And the MCC score of the new resultant algorithm is 0.89.

**Table 3: After applying Grid search generated parameters**

| Models | Recall value | F1-score | MCC value |
|---|---|---|---|
| Random Forest with random parameters | 0.89 | 1.00 | 0.848 |
| Random Forest with Grid Search parameters | 0.90 | 1.00 | 0.89 |

From the above table, we can observe that the best value as compared to the best score of MCC that is 1. The nearest value is generated by the Random Forest algorithm with new parameters that were generated by the Grid search algorithm that is 0.89. Now, from here, we can conclude that we are getting some better results.

## V. CONCLUSION

Credit card fraud detection dataset is used with machine learning algorithms in this experiment to identify which algorithm works better with Credit card fraud detection. In total five algorithms such as SVM, Naïve Bayes, Logistic Regression, KNN, and Random Forest. In which the best score result is given by Random forest and then KNN. As the MCC is used to measure the performance of an algorithm, the best score of MCC is 1 and its values lie between -1 and 1.

Based on MCC measured values Random Forest generated the nearest score of 1 that is 0.848. When the Grid Search algorithm is applied to it and again the model is trained by new parameters the result of Random forest improved a little that the MCC value of Random Forest increased by 0.848 to 0.89 that is again quite near to the best score of MCC that is 1. Based on which we conclude that the best results are given by the Random Forest algorithm in credit card fraud detection.

We can improve the algorithms by combining it with different algorithms and applying new technologies with the algorithms to generate more accurate results with credit card fraud detection. That will help us to reduce the fraud by detecting it at the initial stage. It will reduce the number of losses made by credit card fraudster.

## REFERENCES

1. Shiyang Xuan, Guanjun Liu, Zhenchuan Li, Lutao Zheng, Shuo Wang, Changjun Jiang, Random Forest for Credit Card Fraud Detection,2018 (IEEE).
2. Dejan Varmedja, Mirjana Karanovic, Srdjan Sladojevic, Marko Arsenovic, and Andras Anderla,Credit Card Fraud Detection - Machine Learning methods, Publish in:18th International Symposium INFOTEH-JAHORINA, 20-22 March 2019 (IEEE).
3. Shantanu Rajora, Dong-Lin Li, Chandan Jha, Neha Bharill, Om Prakash Patel, Sudhanshu Joshi, Deepak Puthal, Mukesh Prasad,A Comparative Study of Machine Learning Techniques for Credit Card Fraud Detection Based on Time Variance,2018 (IEEE).
4. Rimpal R. Popat and Jayesh Chaudhary,A Survey on Credit Card Fraud Detection using Machine Learning, 2018 (IEEE), pp. 1120 - 1125
5. Changjun Jiang, Jiahui Song, Guanjun Liu, Member, IEEE, Lutao Zheng, and Wenjing Luan,Credit Card Fraud Detection: A Novel Approach Using Aggregation Strategy and Feedback Mechanism,2018 (IEEE), pp. 2327-4662.
6. Shail Machine, Emad A. Mohamad, Behrouz Far,Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study, 2018(IEEE) computer society, pp.122-125.
7. SamanehSorournejad, Zahra Zojaji, Reza Ebrahimi Atani, Amir Hassan Monadjemi,A Survey of Credit Card Fraud Detection Techniques: Data and Technique Oriented Perspective.
8. Kuldeep Randhawa, Chu Kiong Loo, Manjeevan Seera, Chee Peng Lim, Ashoke K. Nandi,Credit Card Fraud Detection Using AdaBoost and Majority Voting, Published in: IEEE Access on 15 February 2018, vol. no.6, pp. 14277 – 14283.
9. N. Sivakumar, Dr.R. Balasubramanian, Fraud Detection in Credit Card Transactions: Classification, Risks and Prevention Techniques, Published in (IJCSIT) International Journal of Computer Science and Information Technologies, vol no. 6 (2), 2015, pp. 1379-1386.
10. Sai Kiran, Jyoti Guru, Rishabh Kumar, Naveen Kumar, Deepak Katariya, Maheshwar Sharma,Credit card fraud detection using Naïve Bayes model based and KNN classifier, Published in: International Journal of Advance Research, Ideas and Innovations in Technology, Issue no. 3, vol.no. 4, 2018, pp.44-47.
11. Rishi Banerjee, Gabriela Boural, Steven Chen, Mehal Kashyap, Sonia Purohit, Jacob Battipagali,Comparative Analysis of Machine Learning Algorithms through Credit Card Fraud Detection, 2018.
12. Kaithekuzhical Leena Kurien and Dr. Ajeet Chikkamannur,Detection and prediction of credit card fraud transactions using machine learning, Published in: International Journal of engineering science & research technology (IJESRT), 2019, pp. 199-208.
13. Md.Akster Hossain and Mohammed Nazim Uddin,A Differentiate Analysis for Credit Card Fraud Detection, Published in: Int. Conf. on Innovations in Science, Engineering and Technology (ICISET), 27-28 October 2018 (IEEE), pp. 328-333.
14. Suresh K Shirgave, Chetan J. Awati, Rashmi More, Sonam S. Patil,A Review on Credit Card Fraud Detection Using Machine Learning, Published in International journal of Science and Technology Research, vol.no.8, Issue no. 10, October 2019, pp. 1217-1220.
15. S P Mani raj, Aditya Saini, Swarna Deep Sarkar Shadab Ahmed,Credit Card Fraud Detection using Machine Learning and Data Science, Published in: International Journal of Engineering Research & Technology (IJERT), vol.no.8, Issue no. 09, September 2019, pp.110-115.
16. Yashvi Jain, NamrataTiwari, ShripriyaDubey, Sarika Jain, A Comparative Analysis of Various Credit Card Fraud Detection Techniques, Published in: International Journal of Recent Technology and Engineering (IJRTE), vol.no. 7, Issue no. 5S2, January 2019, pp. 402-407.
17. Sonal Mehndiratta and Mr. Kamal Gupta,Credit Card Fraud Detection Techniques: A Review, Published in: International Journal of Computer Science and Mobile Computing, vol.no.8, Issue no.8, August 2019, pp. 43-19.

## AUTHORS PROFILE

**Nayan Uchhana,** pursuing B.E. (CSE) from Indore Institute of Science & Technology Indore. My research interest are in Big Data Analytics, Hadoop, Machine learning, Deep Learning and Artificial Intelligence. I have Published various paper in International and national Journals.

**Ravi Ranjan,** pursuing B.E. (CSE) from Indore Institute of Science & Technology Indore. My research interest is in Machine Learning and Web Development.

**Shashank Sharma,** and I completed my Bachelors of Engineering in Information Technology from IET DAVV, Indore. Currently, I am working as a Software Developer. My research areas include Machine Learning, Artificial Intelligence, Big Data Analytics and Cloud Computing.

**Deepak Agrawal,** and I am Persuing my Ph.D from Sage University, Indore.. Currently, I am working as a Assistant Professor in department of Computer science, IIST, Indore. My research areas include Data structure and algorithm, software testing, data analytics, Machine Learning, Artificial Intelligence, Big Data Analytics and Cloud Computing.

**Anurag Punde,** and I have completed my masters degree from RGPV, Bhopal.. Currently, I am working as a Assistant Professor in department of Computer science, AITR, Indore. My research areas include Cloud Computing, software testing, data analytics, Machine Learning, Artificial Intelligence, Big Data Analytics and Cloud Computing.